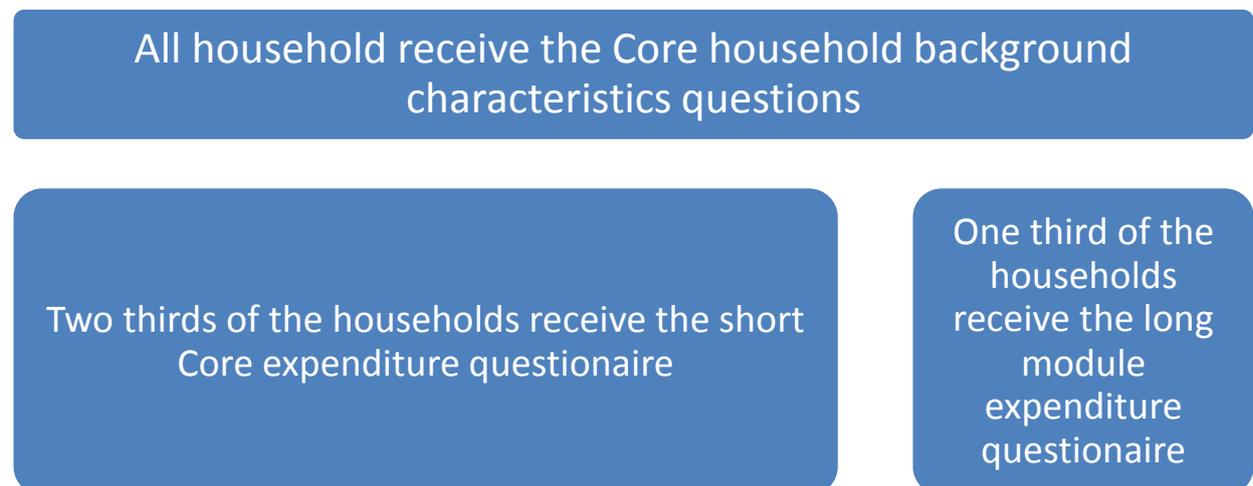


Case B

The Susenas has a Core/Module structure. The Core collects socio-economic data for a large sample of households. The core also includes a short expenditure questionnaire of around 30 questions. The categories in the Core expenditure questionnaire are the same, but the level of aggregation at which the questions are asked is higher (For example, the core asks for expenditure on vegetables, while the Module asks for expenditures on green beans, spinach etc...).

Around one third of the sample receives the Module in addition to the Core. As these households receive a very detailed expenditure question, they are not asked the Core expenditure questions. We thus never have a measure for the short (Core) and the long (Module) expenditure questionnaire for the same household. The Figure below graphs the Core/Module structure



We have provided you with data that includes both the Core and Module households for 1996, 1999 and 2002. We have aggregated the expenditure data from the module to the aggregates as they are asked directly in the Core. For 2002 the Module expenditure data have been discarded. But we have kept the Core expenditure data. Your assignment is to improve upon the estimates you made yesterday.

The large sample of household characteristics is also a stratified random sample. The variable *korweight* is an expansion factor for the household. Thus, for each year;

$$\text{Population size} = \sum_{i=1}^N hmemb_i * korweight_i$$

Where *N* is the number of households in the Core sample. The Module sample, is a random subsample of the Core sample, with a different stratification. For the households that do not appear in the Module, the variable *modweight* is missing.